

Does Sequence Dictate Structure Which Dictates Function?

Jung C. Lee*

BioMolecular Engineering Program and Department of Physics and Chemistry, Milwaukee School of Engineering, Milwaukee, Wisconsin 53202

Received: June 21, 2016; Accepted: July 12, 2016; Published: July 15, 2016

*Corresponding author: Jung C. Lee, Ph.D. Assistant Professor, BioMolecular Engineering Program and Department of Physics and Chemistry, Milwaukee School of Engineering, 1025 N Broadway, Milwaukee, Wisconsin 53202, USA, Tel: +414-277-7316; Fax: +414-277-2878; Email address: lee@msoe.edu

Abstract

Bioinformatics tools and computational methods to predict biomolecular structure from sequence has been and still is in constant development, originally motivated by the Anfinsen's dogma on protein folding. The dogma was the very basis for the development of the extremely widely accepted notion that "sequence dictates structure which dictates function." Nonetheless, the dogma does not support the concept of divergent evolution, the most common form of evolution in nature, but support the concept of convergent evolution, creating several major problems in its application to biomolecular structure prediction. Besides, the dogma ignores homology, the most important requirement for the successful use of comparative sequence analysis, which is the most powerful and most widely used Bioinformatics tool to align homologous sequences not only to infer RNA secondary structures accurately, but also derive evolutionary relationships between diverse organisms. Now is the time to revisit the dogma and throw the ingrained and flawed conventional notion away, followed by adopting a new notion: "Function dictates structure which, in turn, dictates sequence."

Biomolecular structure prediction as a grand challenge in Bioinformatics

Biomolecular structure prediction is one of the grand challenges in Bioinformatics and Computational Biology (1), largely motivated by the work of Christian B. Anfinsen on protein folding. The Anfinsen's Dogma, also known as the thermodynamic hypothesis (2), states that a protein's native three-dimensional (3D) structure is a unique, thermodynamically stable and kinetically accessible global minimum in the Gibbs free energy and is determined solely by its amino acid sequence, first connecting the amino acid sequence to the functional 3D structure and demonstrating a possibility that protein structure could be inferred directly from sequence. Based on his seminal contribution to protein folding, Christian B. Anfinsen was awarded one half of the shared 1972 Nobel Prize in Chemistry. Ever since, the dogma has dominated the area of protein structure prediction, providing the very foundation for the famous notion that sequence dictates structure which, in turn, dictates function. Does the dogma really hold true?

Homology as the key for deriving biomolecular structure

Today's Earth is home to approximately 11 million species (3). Life on Earth evolves from a single common ancestor, or the last universal common ancestor (LUCA) (4). As organisms evolve and diverge conservatively, biosequences also change in response to a set of new environmental constraints, with their respective biological function or homology still maintained throughout (Figure 1A). Homology is all about functional relatedness; it is all or nothing, but not at all equivalent to either sequence similarity or even structural similarity. In fact, homology is the single most important requirement for the success of comparative sequence analysis, the most powerful and most widely used Bioinformatics tool to infer RNA secondary structures and/or derive phylogenetic relationships between diverse organisms on Earth (5). Comparative sequence analysis begins with a compilation and subsequent alignment of multiple homologous biosequences from various different organisms, based on the simple premise that homologous biosequences adopt very similar, if not identical, higher-order structures, regardless of their sequence similarity, in order to maintain their homologous biological function throughout divergent evolution. In contrast, convergent evolution cares only about the accidental evolutionary convergence of sequences or structures, regardless of their shared homology, ending up with multiple exactly the

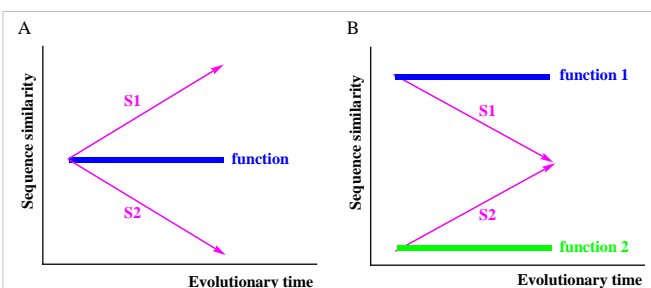


Figure 1: Homology vs sequence similarity. Homology is not equivalent to sequence similarity. (A) Divergent evolution leads to two very different sequences S1 and S2 but maintains its original function and thus homology. (B) Convergent evolution leads to two identical sequences S1 and S2 but maintains no homology.

same sequences or structures, each with a still vastly different biological function (Figure 1B). Without homology, however, a set of proteins with an identical sequence does not necessarily guarantee an identical function. In the same logic, any two proteins with an identical structure, if not homologous, do not possess a similar function. For instance, while human ubiquitin and small ubiquitin-like modifier (SUMO) proteins fold into their native 3D structures whose cores are almost identical, they perform overall opposite biological functions to each other, the former destroying other used proteins and the latter stabilizing other nascent proteins (Figure 2). This strongly suggests that, without homology, sequence-based biomolecular structure prediction will be misleading and a far cry from accurate.

Problems with the Anfinsen's Dogma

Unfortunately, the Anfinsen's dogma itself negates homology or the evolutionary changes of protein sequences and their structures, implying that any change in a protein sequence – natural or erroneous – should disrupt the protein's native 3D structure and subsequently function; it takes into account sequence identify as the sole determinant of biomolecular structure and function. As a consequence, the dogma has long left biologists and geneticists with the “twisted and burning” notion that sequence changes in a protein, if not synonymous, are always deleterious and harmful. In fact, however, homologous biosequences are in a constant flux responding and adjusting to their newly established evolutionary pressures while maintaining their function, as well do their structures. This indicates that

the dogma severely goes against divergent evolution, the most common form of evolution and the one providing the very basis of comparative biosequence analysis, supporting convergent evolution.

Biological robustness as basis for biomolecular structure prediction

Biological robustness is a ubiquitous but fundamental property of a complex biological system to maintain its function against changes, internal or external (6). Under divergent evolution, biosequences and their structures are highly and relatively highly perturbed by changes in environments, respectively, but their function rarely changes, thereby remaining least perturbed and highly conserved. Simply put, divergent evolution makes major room for change in sequence (or *sequence space*), minor room for change in structure (or *structure space*), and nearly no room for change in function (or *function space*) (Figure 3). If evolution is a natural process occurring spontaneously, it will progress into the direction of increasing disorder or entropy, consistent with the second law of thermodynamics. Thus, changes in sequences – exchanges like single nucleotide variations (SNVs), insertions and deletions (indels), and copy number variations (CNVs) – can lead to minor changes in structure, but they will trigger little or no changes in function as a means to maintain biological robustness. This strongly suggests that, in order to maintain biological robustness against environmental perturbations, function should dictate structure which, in turn, should dictate sequence, but not the other way around.

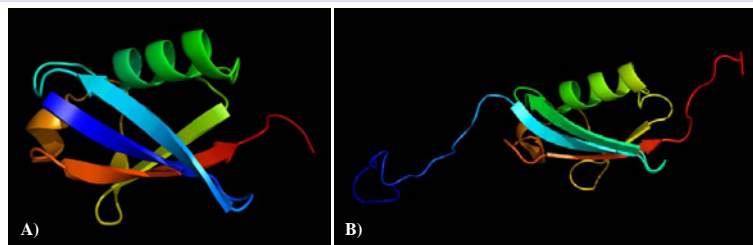


Figure 2: Homology vs structural similarity. Homology is not equivalent to structural similarity. While two different biosequences with no homology can adopt very similar or identical folds, their biological functions are vastly different from each other. (A) The human ubiquitin protein targets other proteins for destruction after use (PDB ID 1UBI). (B) The human small ubiquitin-like modifier 1 (SUMO-1) protein tags other proteins for stabilization after synthesis (PDB ID 1A5R).

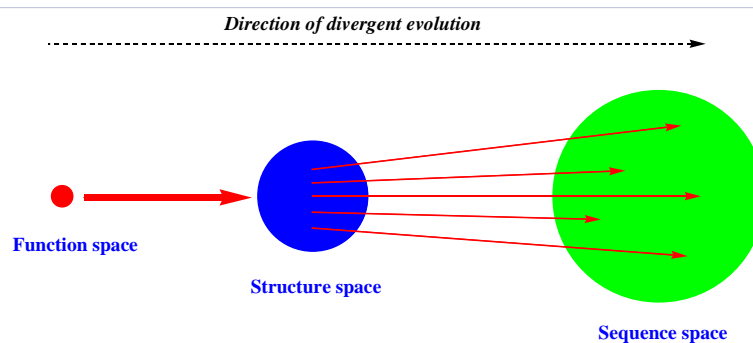


Figure 3: Biological robustness and function-structure-sequence relationship. Under divergent evolution, function dictates structure which, in turn, dictates sequence.

Function dictates structure which, in turn, dictates sequence

Biomolecular structure prediction is still a grand challenge in Bioinformatics and Computational Biology. Besides, biomolecular structure cannot be simply determined solely based on sequence. In particular, both homology and biological robustness must be seriously taken into account in determining biomolecular structure from sequence; they explain systematically and logically how biological functions are so highly conserved despite frequent sequence and structure variations in response to constant environmental changes and other evolutionary pressures. Taken together, it's the time to revisit the Anfinsen's dogma carefully to drop the deeply ingrained and misleading conventional notion of "sequence dictates structure which dictates function", and subsequently adopt a new notion: "Function dictates structure which, in turn, dictates sequence."

Acknowledgements

The work was supported in part by Faculty Development Grant awarded to the author from Milwaukee School of Engineering.

Source of Support: Faculty Development Grant, Milwaukee School of Engineering

References

1. Yang Zhang, Progress and challenges in protein structure prediction. *Curr Opin Struct Biol.* 2008;18(3): 342-348. doi: 10.1016/j.sbi.2008.02.004
2. Anfinsen CB. Principles that govern the folding of protein chains. *Science.* 1973;181(4096): 223-230.
3. Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on Earth and in the ocean?. *PLoS Biol.* 2011; 9(8):e1001127. doi:10.1371/journal.pbio.1001127
4. Gouy M, Chausson M. Evolutionary biology: ancient bacteria liked it hot. *Nature.* 2008; 451(7179):635-636. doi: 10.1038/451635a.
5. Gutell RR, Lee JC, Cannone JJ. The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol.* 2002;12(3): 301-310.
6. Kitano H. Biological robustness. *Nat Rev Genet.* 2004;5(11): 826-837. doi:10.1038/nrg1471.